

**Mapping Processes in Motion:
Visualizing the ICPSR Data Pipeline**

Cole Whiteman
Inter-university Consortium for Political and Social Research (ICPSR)
P.O. Box 1248, Ann Arbor, MI 48106
734-615-7935
colew@umich.edu

ABSTRACT

In this case study we look at one repository's "process mapping" approach to coping with a large and complex body of technical procedures that change frequently and tend to get out of synchronization with each other. ICPSR has invested effort over the past 3 years to map its data pipeline process, in order to increase shared understanding among staff about how the process works, and to guide multiple process improvements.

SCENARIO

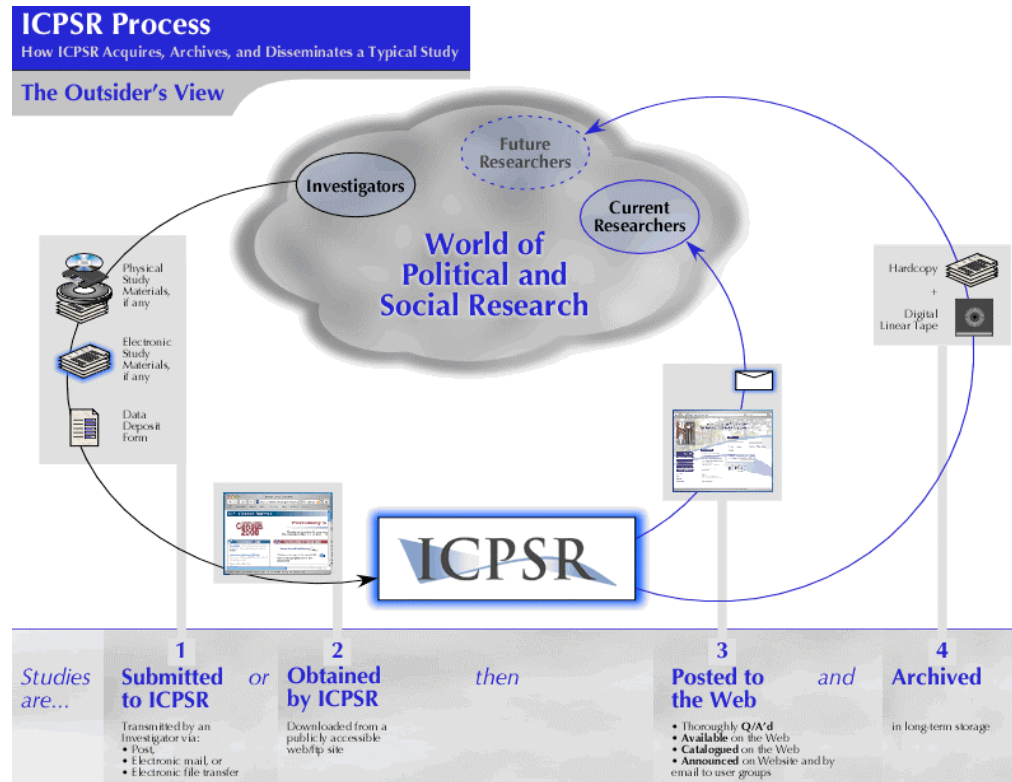
The Case Study: ICPSR

<i>The Setting</i>	<p>The setting for this case study is ICPSR, the Inter-university Consortium for Political and Social Research, which has been hosted since its inception in the early 1960's by the University of Michigan's Institute for Social Research.</p> <p>ICPSR operates a suite of social science data archives as a service to the research community. What comes into the ICPSR are social science data collections, appearing in a variety of formats and states of assembly; what goes out are standardized, validated, fully documented versions of these collections, posted on the web for today's researchers and archived for researchers of the future.</p>
<i>The Pipeline</i>	<p>How "what comes in" becomes "what goes out" is the product of a data acquisition and processing pipeline that involves many groups, many people, many procedural steps, and many technical tools. The pipeline components have been assembled over ICPSR's 40-plus year history by people dedicated to improvising ways to handle the wide variety of input that the organization receives. Some of the developers of these procedures and tools are still on staff, and they serve as sources of lore about how these components work and why they are the way they are.</p>
<i>The Problem</i>	<p>Over the past decade there arose within ICPSR an appreciation that, while many of the pipeline components have adapted to numerous changes in technology and in user expectations, others have been less responsive; and that this was at least in part because there was no straightforward way for any one person to get a grasp on how the whole pipeline worked, at either a summary or a detailed level.</p> <p>There had been several attempts over the years to document the pipeline operation, but the results had been less than satisfactory for various reasons: the output documents were incomplete, or lacking in detail, or incoherent, or otherwise unclear. Some documentation attempts were considered "partisan", having been written from one or another group's perspective, without regard for variations in procedures among the groups.</p>
<i>The Solution</i>	<p>Three years ago, ICPSR began to respond by investing in a focused effort to document its pipeline process — systematically, coherently, comprehensively, spanning the entire pipeline from beginning to end, integrating all relevant perspectives, and with results made easily available and comprehensible to all involved staff.</p>

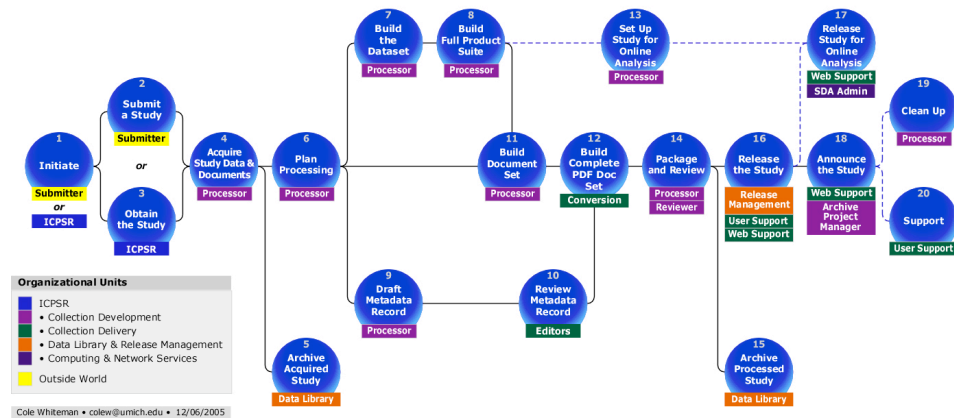
We spent a summer creating an initial set of hierarchical diagrams to depict the pipeline process, with a one-page high-level summary at the top of the hierarchy and a detailed map measuring 3 feet high by 28 feet long at the bottom. We posted these on a corridor wall for staff and visitors to see.

Samples

Here is our diagram of how the process looks to the outside world. The ICPSR operation is an opaque block that receives “raw” studies and delivers web-posted and archived studies.

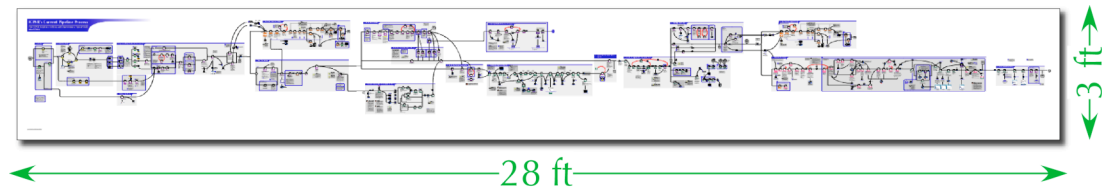


The pipeline behind the scenes, at the top level, looks like this:



Each bubble represents a major step in the pipeline, and has a detailed diagram behind it.

The composite of the detailed diagrams is the one that spans the corridor wall:



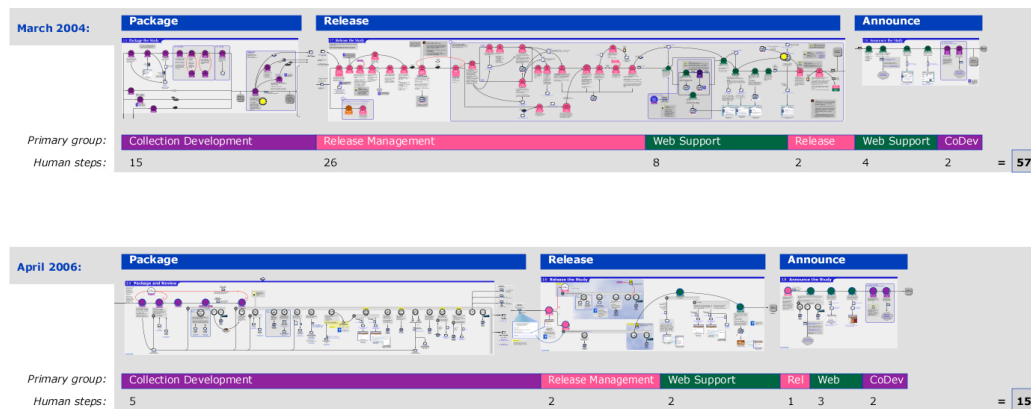
Benefits

This process mapping exercise brought to light a body of distributed and submerged lore about how the process actually works, plus staff desires for how it might work or should work.

The process mapping led directly to a coordinated process improvement effort, with a rotating team devoted to gathering ideas and developing them into proposals. This in turn has yielded a stream of pipeline enhancements, some conservative, some radical. We are now seeing the results accumulate in the form of increased throughput with less effort.

The process map itself has become clearer and more streamlined; we've been periodically updating it to keep up with pipeline automation and other improvements. The map has also been useful as a source for a number of derivative documents we've created for planning and presentation. For example: here's one to illustrate "before and after" a wave of automation on the tail end of the pipeline:

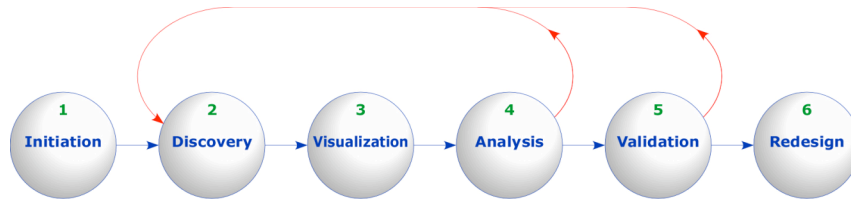
ICPSR Pipeline Process: Package – Release – Announce Steps: Comparing Old and New



As another spinoff from the process mapping exercise, we are now overhauling the internal technical procedure documents on our Intranet web site, using our high-level process diagram as an organizing principle and as a visual table of contents.

The Mechanics of Process Mapping

How does an organization go about constructing a solid, useful process diagram? We find that this kind of effort involves six activities, which can overlap somewhat but commence in roughly the following order:



1. Initiation

This is where we plan the diagramming effort. We clarify:

- our essential goal;
- our intended audience;
- the scope of the process(es) we're going to describe;
- what tools we'll use to construct the diagrams;
- what format(s) we'll render it in for distribution;
- who will play what role in the effort; and
- any constraints on time or resources.

2. Discovery

This is where we gather information about the process we're describing. It includes talking to people, reading existing documents, and navigating through existing information systems. The "talking to people" can appear informal but is structured around:

- *Interviews*, in which we collect people's stories about how their part of the process works (which we consider authoritative), and how other parts of the process work (which we consider very useful but not authoritative).
- *Reviews*, in which we gather people's reactions to others' stories, and to our visual rendition of these stories, as a way to detect inconsistencies.
- *Discussions*, in which we pull together two or more people who have given conflicting (or apparently conflicting) tales of how something works, so we can resolve the inconsistencies.

3. Visualization

This is where we actually sit down and draw something. We might do early sketches by hand, but pretty quickly we turn to a computer and draft a process diagram using illustration software. Depending on the audience, we might be creating a large diagram to hang on a wall, or a hyperlinked set of small diagrams to post on the Web.

For graphics software, we believe that most any of the major illustration packages on the market would serve well. For the ICPSR process mapping effort we chose *Canvas* by ACD Systems because it is cross-platform (Windows and Mac — ICPSR has staff of both persuasions); it integrates bitmap images and vector object graphics, so we can use one piece of software instead of two; it has a library of common graphic symbols, plus a rich tool set for designing new graphics so we can create icons for locally familiar process objects; it can handle large illustration sizes (for hanging posters on walls); and it exports/imports well with PowerPoint and other graphic applications.

As an alternative approach, there are software packages on the market that are specifically devoted to project management or process design; however, we chose to go with illustration software out of a sense that our project would be more about clear visual explanation of a process rather than (say) computation of a critical path.

4. Analysis

This is where we apply a programmer's discipline to the diagram, checking its visual syntax as though it were a body of executable code. We look at:

- *Terminology control*. We look for violations of consistent terminology, as when the same term is used to mean two different things, or two terms are used to mean the same thing. As an example at ICPSR of the first problem, we found that different groups

would speak of entering data into or extracting data from "The Database" but they were (unknowingly) referring to different databases. As another example, we found that staff used the term "restricted data" to mean three quite different things, depending on the context. As an example at ICPSR of the second problem, we found that the staff were (collectively) using four very different pieces of jargon to mean the same thing: the record of metadata they were assembling for a data collection.

- *Object management.* We look for undocumented or dangling data, materials, storage; objects that are created but don't appear to be used; objects that appear out of nowhere; and objects that appear to be mysteriously moved, copied, deleted, renamed, or otherwise modified. Such anomalies may merely indicate that the diagram is not yet accurate, and some more discovery is called for; but often they reveal that some aspect of the process is poorly understood by the organization, if not outright broken. At ICPSR we found several examples of data, documents, or database elements that were created and handed off to another group, but then never used.
- *Object ownership.* We look for orphaned processes, which have no clear owner; and contested processes, over which two or more people or groups claim ownership. At ICPSR we saw numerous examples of this around quality control. Nearly every group involved in the pipeline spoke of the valuable checks that they performed, and of the other checks that they believed were being done by other groups. However, but on examination of quality control across the entire pipeline it turned out that certain checks were actually being done several times (redundantly) and others were not being done at all.

5. Validation

This is where we track a sample of actual work going through the process, and update the diagram accordingly. If the Analysis step is analogous to program syntax checking at compile time, this Validation step is analogous to program checking at run time. Does the program actually execute? Does an actual piece of work follow one of the paths shown on the diagram, or does it enter uncharted territory?

At ICPSR we followed some data collections through the pipeline. We watched as they were handed off from one group to another; we tracked them with sticky notes on the diagram on the corridor wall; and we kept a detailed online diary to record observations.

We found that although the data collections proceeded through the pipeline mostly as described, there were several points of surprise, which revealed situations that we hadn't previously heard about. We followed up with more investigation and then incorporated additional paths and clarification into the diagram.

6. Redesign

This is where we propose changes to the process we've diagrammed.

In the course of constructing the diagram, we find that many of the conversations about how the process currently works are peppered with stories about how the process doesn't work, might work in the future, or should be made to work.

We capture these plans, suggestions, and wishes; and we incorporate them into the diagram or prepare separate diagrams, so as to inform focused discussions on redesigning the process.

Stability and Change

So does this case study support the colloquium hypothesis that "What we do remains the same, how we do it changes?"

We could debate this as stated, but allow me to respond with an alternative formulation that is less black-and-white:

- ICPSR's collection of activities form a spectrum between "What we do" and "How we do it".
- On one side of the spectrum, there is an essence to "what we do" that has indeed remained the same. ICPSR's core service is still preservation and dissemination of social science data collections; we still ingest data, evaluate it, process it, archive it, and publish it.
- On the other side of the spectrum, there are endless details of "how we do it", and they will change endlessly in response to evolution in the services we offer, the technology we use, the organization we belong to, and the communities we serve.
- There is a grey area in the middle, which can look like "what" or "how" according to one's point of view. For example, our webmaster might think of our activity of publishing data on the Web to be what we do; long-time staff in other departments might say that data dissemination is what we do, and web publishing is how we do it currently.

Where is process mapping on that spectrum? I will suggest that one of ICPSR's organizational activities has always been "striving to understand how we do what we do", and that process mapping gives us a new way to do it effectively.

NEW SKILLS

The sort of process mapping effort we describe here requires the organization to bring together the process content experts and formal process owners (who may or may not be the same people) with the skill sets of five roles:

- an *Investigator*, who contributes skills in effective interviewing, unbiased observation, careful recording, and logical deduction;
- a *Process Designer*, who contributes skills in understanding the logic of complex procedural sequences in a technical organization and making them more efficient and effective;
- a *Technical Communicator*, who contributes skills in clear writing and logical presentation;
- a *Graphic Designer*, who contributes skills in rendering ideas visually by hand or, these days, using software; and
- a *Web Publisher*, who contributes skills in crafting documents for the web, and putting them in place on a web site.

And, as with any project, these people could use the support of:

- a *Sponsor*, who provides the organizational resources for the effort;
- a *Project Manager*, who provides coordination and tracking; and perhaps even
- an *Online Document Librarian*, who organizes and maintains the project's collection of graphic and text documents in an orderly file system.
-

In practice, you may be able to find people who can fill more than one role and contribute more than one skill set.

DISCUSSION QUESTIONS

Does your repository organization have to cope with a large and complex body of technical procedures that change frequently and get out of synchronization with each other? Is this kind of problem unique to ICPSR or common among repository organizations?

Does your organization have experience with attempting to address this problem in any kind of focused way?

Would the process mapping approach described here, or some variation on it, be helpful in your organization? What would have to be true to ensure that it would be a helpful exercise, rather than a time-consuming distraction?

Would it be easy or hard to assemble a process information design team with all required skills represented, and with the time available to devote to such an effort? Which of these skill needs would you find it easiest to fill? Hardest?

Might you find some colleagues reluctant to cooperate with a process documentation effort? What reasons might they have?

How can we learn to create and maintain process documentation rapidly enough to capture the state of a process before it changes again?
